



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **06290251 A**(43) Date of publication of application: **18.10.94**

(51) Int. Cl.

G06F 15/62
G06K 9/20
H04N 1/21

(21) Application number: **05073348**(22) Date of filing: **31.03.93**(71) Applicant: **HITACHI LTD**

(72) Inventor: **KODAMA KAZUYUKI**
MURAKAMI TATSUYA
IKUNO YOSHITSUGU
HATANO HIDEKAZU
FUJINAWA MASAOKI

(54) **DOCUMENT IMAGE PROCESSING METHOD OF ELECTRONIC FILE DEVICE**

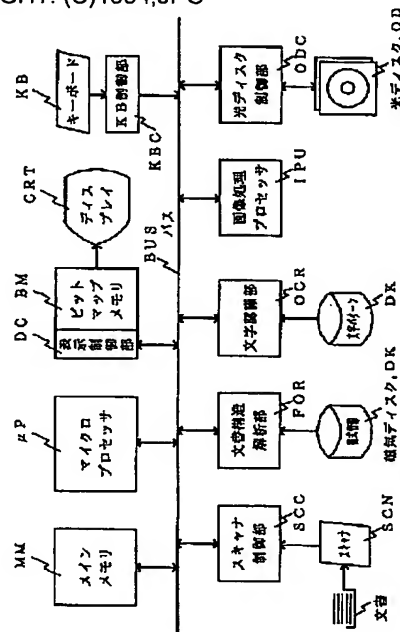
the area is hidden.

COPYRIGHT: (C)1994,JPO

(57) Abstract:

PURPOSE: To display image data while masking only a secret item entry area even if the document format is different by making the secret item entry area of the image data according to coordinate values and displaying the data other than the area in the state of the original image.

CONSTITUTION: A main memory MM stores the program of a microprocessor and also stores image data, etc., outputted from a scanner SCN temporarily. An optical disk OD stores with image data which are inputted from the scanner SCN and compressed by an image processing processor IPU and document indexes converted into text data by a character recognition part OCR, i.e., key words for document retrieval correlatingly. For registration, the document structure is analyzed to extract the entry item of a secret item, whose coordinate values are stored on the optical disk OD together with the image data. Then, when a document image is used, an image mask pattern corresponding to position information on the secret item entry area is generated to display or print a document image wherein,



(51)Int.Cl.⁵

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 15/62

3 3 0 A

8125-5L

G 0 6 K 9/20

3 4 0 C

H 0 4 N 1/21

7232-5C

審査請求 未請求 請求項の数 4 O L (全 13 頁)

(21)出願番号 特願平5-73348

(22)出願日 平成 5年(1993) 3月31日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目 6 番地

(72)発明者 児玉 和行

東京都国分寺市東恋ヶ窪 1 丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 村上 達也

東京都国分寺市東恋ヶ窪 1 丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 生野 尚嗣

東京都国分寺市東恋ヶ窪 1 丁目280番地

株式会社日立製作所中央研究所内

(74)代理人 弁理士 小川 勝男

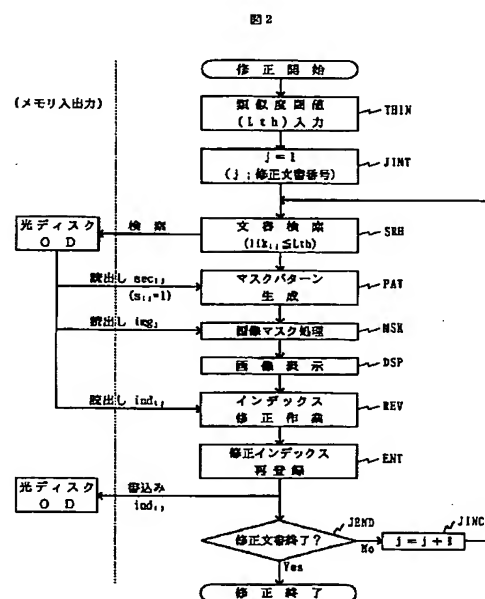
最終頁に続く

(54)【発明の名称】 電子ファイル装置の文書画像処理方法

(57)【要約】

【目的】 文書イメージからインデックスを認識することにより、大量の文書を自動的に登録する電子ファイリング装置に関し、特に守秘項目が存在する文書を対象とし、守秘項目をマスクした画像表示などの出力を行なうための文書画像処理方法を提供する。

【構成】 登録においては、あらかじめ作成した書式情報を利用して文書構造を解析し、守秘項目の記載領域を抽出して、その座標値を、画像データとともに光ディスクなどの大容量のメモリへ格納する。文書イメージを使用する際には、守秘項目領域に対応した画像マスクパターンを生成し出力する。また、インデックス修正に際しては、守秘項目領域の座標値を読み出し、該領域に対応した画像マスクパターンを生成し、画像データと該マスクパターンとの論理演算をした後、ディスプレイなどに出力する。



【特許請求の範囲】

【請求項1】紙面上の文書イメージを光電変換するスキャナと、

該変換されたイメージを一時記憶する第一のメモリと、
該メモリに記憶されたイメージ情報から文書の書式を同定して、インデックス記載領域および守秘項目記載領域を抽出する文書構造解析部と、

上記インデックス記載領域のイメージ情報を第一のメモリから読み出し、文字パターンを認識してテキストコードに変換する文字認識部と、

多数の文書イメージとそのインデックスを格納する大容量の第二のメモリとを具備し、表形式の文書を登録し、登録した文書を出力する電子ファイル装置であって、

(a)上記第一のメモリに記憶された文書イメージと、文書構造解析で抽出した守秘項目記載領域の位置情報と、文字認識したインデックスのテキストコードを関連付けて第二のメモリへ格納し、

(b)格納した文書イメージを使用するに際し、上記守秘項目記載領域の位置情報に対応した領域を覆い隠した文書イメージの表示あるいは印刷をすることを特徴とする電子ファイル装置の文書画像処理方法。

【請求項2】電子ファイル装置使用者が入力するパスワードにより、格納した文書イメージの全領域を表示あるいは印刷する動作モードと、守秘項目記載領域を覆い隠した文書イメージを表示あるいは印刷する動作モードとを切り替えることを特徴とする請求項1記載の電子ファイル装置の文書画像処理方法。

【請求項3】紙面上の文書イメージを光電変換するスキャナと、

該変換されたイメージを一時記憶する第一のメモリと、
該メモリに記憶されたイメージ情報から文書の書式を同定して、インデックス記載領域および守秘項目記載領域を抽出する文書構造解析部と、

上記インデックス記載領域のイメージ情報を第一のメモリから読み出し、文字パターンを認識してテキストコードに変換するとともに、該パターン認識の類似度を出力する文字認識部と、

多数の文書イメージとそのインデックスを格納する大容量の第二のメモリとを具備し、表形式の文書を登録し、登録した文書を出力する電子ファイル装置であって、

(a)上記第一のメモリに記憶された文書イメージと、文書構造解析で抽出したインデックス記載領域および守秘項目記載領域の位置情報と、文字認識したインデックスのテキストコードと類似度を関連付けて第二のメモリへ格納し、

(b)格納した文書イメージを用いてインデックスを修正するに際し、類似度があらかじめ設定した値以下の文書イメージを第二のメモリから検索して読み出し、上記守秘項目記載領域の位置情報に対応した領域を覆い隠した文書イメージの表示あるいは印刷を行なう、ことを特徴

とする電子ファイル装置の文書画像処理方法。

【請求項4】請求項3記載の電子ファイル装置において、

(a)①認識の類似度があらかじめ設定した閾値以上の場合、すなわち認識の信頼度が高い場合には、認識したインデックスのテキストコードを第二のメモリへ格納し、

②認識の類似度が閾値以下の場合には、第一のメモリに記憶された文書イメージと、文書構造解析で抽出したインデックス記載領域および守秘項目記載領域の位置情報と、認識したインデックスのテキストコードを関連付けて第二のメモリへ格納し、

(b)格納した文書イメージを用いてインデックスを修正するに際し、上記守秘項目記載領域の位置情報に対応した領域を覆い隠した文書イメージの表示あるいは印刷を行なう、ことを特徴とする電子ファイル装置の文書画像処理方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、文書イメージからインデックスを認識することにより、大量の文書を自動的に登録する電子ファイル装置に関する。特に守秘項目が存在する文書を対象とし、誤認識したインデックスを修正する上で、守秘項目をマスクした画像表示をするに好適な文書画像処理方法に関する。

【0002】

【従来の技術】文書自動登録を目的とした従来技術として、例えば、特開昭61-59568号公報記載の「文書理解方式」では、文書画像から枠部分を抽出して書式情報を作成し、書式情報を用いて文書上の文字を読み取る方式が開示されている。また、特開昭60-101681号公報記載の「文字読取方式」では、文字記入欄毎に記述した書式情報を利用して認識方式を制御することにより、多様な字種の存在する文書の認識精度を高める技術が開示されている。

【0003】また、機密保護に関しては、例えば、特開昭61-285578号公報記載の「ファリング方式」では、文書画像全面に網かけ等特定のパターンを重畳して出力する方式が開示されている。

【0004】

【発明が解決しようとする課題】しかしながら、自動登録した文書を精度良く検索するために、認識結果を修正しておく必要があり、この修正作業は、登録文書数の増加、文書書式の多様化にともない、文書管理部門以外で行なう場合が多くなっている。ここに、多様な書式を有する文書を対象として、インデックス記載領域の画像は鮮明に、部外秘項目記載領域の画像はマスクして（覆い隠して）表示する必要が新たに生じている。

【0005】本発明は、上記事情に鑑みてなされたもので、その目的とするところは、文書イメージからインデックスを認識することにより、大量の文書を自動的に登

録する電子ファイル装置に関し、特に守秘項目が存在する文書を対象として、守秘項目記載領域をマスクして出力する文書画像処理方法を提供することにある。さらに、誤認識したインデックスを修正する上で、守秘項目領域のみをマスクして出力する文書画像処理方法を提供することにある。

【0006】

【課題を解決するための手段】本発明による文書画像処理方法は、帳票に代表される表形式の文書を電子ファイル装置へ自動的に登録する方法、および登録した文書を

表示する方法ものである。

【0007】具体的には、①紙面上の文書イメージを光電変換するスキャナと、②該変換されたイメージを一時記憶する第一のメモリと、③該メモリに記憶されたイメージ情報から文書の書式を同定して、インデックス記載領域および守秘項目記載領域を抽出する文書構造解析部と、④上記インデックス記載領域のイメージ情報を第一のメモリから読み出し、文字パターンを認識してテキストコードに変換するとともに、該パターン認識の類似度を出力する文字認識部と、⑤多数の文書イメージとそのインデックスを格納する大容量の第二のメモリとを具備した電子ファイル装置であって、(a)上記第一のメモリに記憶された文書イメージと、文書構造解析で抽出したインデックス記載領域および守秘項目記載領域の位置情報と、文字認識したインデックスのテキストコード（と類似度）を関連付けて第二のメモリへ格納し、(b)格納した文書イメージを使用するに際し、守秘項目記載領域の位置情報に対応した画像マスクパターンを生成し、該領域を覆い隠した文書イメージの表示あるいは印刷を行なう、または、(b)'格納した文書イメージを用いてインデックスを修正するに際し、類似度があらかじめ設定した値以下の文書データを第二のメモリから検索して読み出し、守秘項目記載領域の位置情報に対応した画像マスクパターンを生成し、該領域を覆い隠した文書イメージの表示あるいは印刷を行なう、ことを特徴とする。

【0008】

【作用】本発明による文書画像処理方法の代表的な作用は、以下の通りである。まず、登録においては、あらかじめ作成した書式情報を利用して文書構造を解析し、守秘項目の記載領域を抽出して、その座標値を、画像データとともに第二メモリへ格納する。

【0009】次に、文書イメージの使用に際しては、守秘項目記載領域の位置情報に対応した画像マスクパターンを生成し、その領域を覆い隠した文書イメージの表示あるいは印刷を行なう。

【0010】また、インデックス修正のための表示では、第二メモリへ格納された守秘項目記載領域の座標値を読み出し、該領域に対応した画像マスクパターンを生成し、画像データと該マスクパターンとの論理演算をした後、ディスプレイに表示する。この方法により、登録

対象とする文書の書式が相違しても、守秘領域のみを覆い隠した画像の表示が可能になる。

【0011】

【実施例】以下、本発明の実施例を図面により詳細に説明する。まず、図8に、電子ファイル装置（電子ファイリング装置ともいう）への自動登録の対象とする文書の書式例を示す。同図は報酬明細書の例で、罫線Lx、Ly等で囲まれた領域F1、F2、F4、およびFnには、文書のインデックス、換言すれば文書検索のキーワードになる項目f1、f2、f4、およびfnが記載されている。具体的には、f1は「報酬明細書平成5年3月分」等の日付、f2は被報酬者の個人識別番号、f4は従事日数、fnは合計金額である。領域F3には、報酬支払関係者以外には見せてはならない守秘項目f3（被報酬者氏名）が記載されている。また、本報酬明細書には、発行事業所毎に領域Fの位置および大きさが少しずつ相違している、項目fの書体が異なるという特徴がある。

【0012】次に、図9を用いて、本発明が適用される文書自動登録機能を有する電子ファイリング装置の構成を説明する。同図中、SCNは、図8で示した書式の文書を1インチ当たり200または400ドットの画素数で光学的に読み取り、1画素当たり1ビットに2値化して画像を出力するイメージスキャナを示す。SCCは、後述するマイクロプロセッサμPの命令により、スキャナSCNとバスとを接続する制御部を示す。

【0013】MMはメインメモリを示し、マイクロプロセッサのプログラムを蓄え、スキャナから出力された画像データ等を一時的に蓄える。MMは4MB（メガバイト）の半導体メモリで構成されている。

【0014】FORは、文書構造解析部を示し、スキャナから入力した画像データの罫線の長さ、始点座標等の特徴を分析し、磁気ディスクDKにあらかじめ蓄積した書式情報と照合することにより入力文書の書式を同定し、インデックス記載領域の座標値を求める。OCRは、文字認識部を示し、上記FORで求めたインデックス記載領域の座標値に基づいて画像データを切り出し、その領域内の文字パターンを認識してテキストデータに変換する。

【0015】IPUは、画像処理プロセッサを示し、①メインメモリMMから読み出された2値画像を、MMR符号化方式を用いて圧縮し、光ディスクに転送する、②逆に、圧縮した画像を伸張し、表示用ビットマップメモリへ送出する、③画像を部分的にマスクする（覆い隠す）パターンを生成する、④伸張した画像データとマスクパターンとのオア演算をする、といった画像処理機能を有する。

【0016】ODは、1枚当たり600MBの記憶容量を持つ光ディスクとその駆動装置を示し、①スキャナSCNから入力され、画像処理プロセッサIPUで圧縮さ

れた画像データと、②文字認識部OCRでテキストデータに変換された文書インデックス、すなわち文書検索用のキーワードを関連付けて蓄積する。ODCは、マイクロプロセッサ μP の命令に応じて、光ディスクODの読み出し／書き込み制御を行なう光ディスク制御部であり、最大100枚の光ディスクを制御することができる。なお、スキャナSCNで読み取った文書画像データは効率良く圧縮されるので、1枚の光ディスクに3万枚程度の文書を登録することが可能である。

【0017】マイクロプロセッサ μP は、キーボードKBから入力されるユーザの処理要求に応じて、スキャナSCN、文書構造解析部FOR、文字認識部OCR、画像処理プロセッサIPU、光ディスクOD等の動作を制御する。BUSは、32ビット幅のデータバスである。CRTは、装置の操作ガイダンス、文書インデックス、および文書画像を表示するディスプレイである。BMは、ビットマップメモリで、CRTに表示する画像データが蓄えられる。DCは、マイクロプロセッサ μP の表示データと画像処理プロセッサIPUから出力される画像データとを合成してビットマップメモリBMへ書き込む表示制御部を示す。KBは装置への処理要求、検索条件等を入力するためのキーボード、KBCはキーボード制御部である。

【0018】次に、文書登録時に、光ディスクの所定のセクタに書込まれるキーワード情報の構成を図10に示す。図は、1文書当たりのキーワード情報を示したもので、情報の長さKは512B（バイト）である。文書のインデックス、すなわちキーワードは、いくつかの項目（項目K1、項目K2、…、項目Kn）について付与することができる。これらの項目の例としては、報酬明細書の場合には、日付、個人番号、従事日数、合計金額等が挙げられる。キーワード構成情報フィールドPには、項目数、各項目名、各項目のデータ長が書かれ、各項目フィールドQには、文字認識部OCRの出力であるキーワードが書込まれる。光ディスク管理情報フィールドRには、文書画像データが蓄積された番地と画像データ長が書かれる。

【0019】本発明による文書画像自動登録、修正表示、および検索表示方法を図1～図7を用いて説明する。特徴は、①登録においては、文書構造解析により求めた守秘項目記載領域の座標値を、画像データおよび文字認識部で認識したインデックスとともに光ディスクに蓄積し、②修正・検索表示では、光ディスクから読み出した画像の守秘項目記載領域をマスクして表示することにより、守秘部分の適正な表示を実現したものである。以下、登録、修正、および検索方法毎に詳細に説明する。

【0020】（1）文書自動登録方法

まず、図6に、書式情報を示す。書式情報は、入力された文書の書式を同定し、各項目が記載された領域の座標

値等を求める上で必要である。図は、書式を定量的に記述する方法を示したもので、文書インデックスが記載された領域、および守秘項目が記載された領域を (X_{ij}, Y_{ij}) 、 ΔX_{ij} 、 ΔY_{ij} で、各項目の属性を a_{ij} 、 s_{ij} で記述する。ここに、

(X_{ij}, Y_{ij}) : 領域F iを形成する罫線の始点座標の値
 ΔX_{ij} : 領域F iを形成する罫線のX方向の長さ
 ΔY_{ij} : 領域F iを形成する罫線のY方向の長さ
 a_{ij} : 領域F iに記載された項目f iの文字の

種類が

機械書き数字のとき、 $a_{ij}=0$

機械書き漢字のとき、 $a_{ij}=1$

手書き数字のとき、 $a_{ij}=2$

手書き漢字のとき、 $a_{ij}=3$

（機械とは、ワードプロセッサ等を言う）

s_{ij} : 領域F iに記載された項目f iが

守秘項目でないとき、 $s_{ij}=0$

守秘項目のとき、 $s_{ij}=1$

である。なお、添字jは、書式番号を表わす。これは、報酬明細書の発行事業所毎に書式、すなわち領域F iの位置および大きさ、項目f iの文字の書体が相違するためである。

【0021】このようにして、図7に示すように、いくつかの書式（書式1～m）に対して書式情報をあらかじめ作成し、書式情報ファイルとして磁気ディスクに用意しておく。

【0022】次に、図1により、文書画像の登録方法を説明する。同図で、キーボードKBから「文書登録」要求が入力されると、

①手順JINTで、マイクロプロセッサ μP は、入力文書番号jを1に設定し、次の手順SCNでスキャナを起動する。スキャナは、手順IMGで文書画像を光学的に読み取り、2値化する。 μP は、画像データimgjをメインメモリMM上のM1番地に書き込む。

【0023】②次に、手順FORで、文書構造解析部は、画像データimgjの罫線の長さ、始点座標等の特徴を分析し、磁気ディスクDKにあらかじめ用意した書式情報TFORと照合することにより入力文書の書式を同定し、該当した書式mの項目記載領域の座標データ (X_{im}, Y_{im}) 、 ΔX_{im} 、および ΔY_{im} （これらを総称してsecijと言う）と、項目の属性データaim、sim（それぞれ a_{ij} 、 s_{ij} と言う）をメインメモリのM2番地に転送する。

【0024】書式同定の具体的な方法は、

a) 入力画像imgjに存在する連続した黒画素を追跡し、閉じた罫線枠を抽出する。

b) 該罫線枠の始点座標 (X_k, Y_k) と、長さ ΔX_k 、 ΔY_k を求める。

c) 書式情報ファイルから $((X_{im}, Y_{im}), \Delta X_{im}, \Delta Y_{im})$ を読み出す。

d) 各書式ごとに、 $((X_k, Y_k), \Delta X_k, \Delta Y_k)$ と $* [0025]$
 $((X_{im}, Y_{im}), \Delta X_{im}, \Delta Y_{im})$ を対応付ける。すな $[数1]$
 わち、すべての領域 i に対して $*$

$$\min_{k(k=1 \sim n)} [(X_{im} - X_k)^2 + (Y_{im} - Y_k)^2] \quad \dots (数1)$$

となる k を求め、領域 k の $((X_k, Y_k), \Delta X_k, \Delta Y_k)$ $\times [0026]$
 を $((X_i, Y_i), \Delta X_i, \Delta Y_i)$ に置き換える。 $[数2]$

e) 次に、各書式に対して、 \times

$$\Delta N_m = \sum_{i=1}^n [(X_{im} - X_i)^2 + (Y_{im} - Y_i)^2] \quad \dots (数2)$$

[0027]

★ ★ [数3]

$$\Delta L_m = \sum_{i=1}^n [(\Delta X_{im} - \Delta X_i)^2 + (\Delta Y_{im} - \Delta Y_i)^2] \quad \dots (数3)$$

[0028]

☆ ☆ [数4]

$$\Delta P_m = \Delta N_m + \Delta L_m \quad \dots (数4)$$

を計算する。

f) 数4の ΔP_m を最小とする書式 m が、入力された文書の書式と同一である。

[0029] g) 書式 m の $((X_{im}, Y_{im}), \Delta X_{im}, \Delta Y_{im})$ 、 a_{im} 、および s_{im} を書式情報ファイルから読み出し、メインメモリへ転送する。

[0030] ③次に、手順SECで、文字認識部は、 $s_{ij} = 0$ (領域 i が守秘項目でない) ならば、上記FORにより求めたインデックス記載領域の座標データ sec_{ij} で規定される範囲の画像データをメインメモリのM1番地から読み出し (画像切り出し)、手順OCRで、該画像中の文字パターンを認識してテキストデータ ind_{ij} に変換する (文字認識)。OCRでは、同時に、パ

ターン照合の類似度 lik_{ij} も出力する。

[0031] なお、文字認識では、項目の属性データ a_{ij} に基づいて、最適な認識方式を選択駆動することにより、認識精度の向上を図ることが可能である。

[0032] ④手順FILでは、画像データ img_j をメインメモリのM1番地から読み出し、画像処理プロセッサIPUで圧縮した後、光ディスクODに蓄積 (ファイレインク) する。各項目記載領域の座標データ sec_{ij} 、守秘項目であるか否かの情報 s_{ij} も画像データと同様に光ディスクに蓄積する。上述OCRの出力であるインデックステキストデータ ind_{ij} 、および類似度 lik_{ij} は、光ディスクのキーワードフィールド (図10で $K1, K2, \dots, Kn$) に書き込む。

[0033] ⑤なお、手順INT、IEND、およびINCは、一文書について所定数の領域の画像切り出し、および文字認識処理を繰り返して行なうことを示 \blacklozenge

$$lik_{ij} \leq Lth$$

なる文書を検索する。

[0038] ③手順PATでは、キーワード情報の光ディスク管理情報フィールドRを参照することにより、s

◆す。また、手順JINT、JEND、およびJINCは、スキヤナにセットされたすべての文書に対して、SCNからFILまでの一連の登録処理を繰り返して行なうことを示す。

[0034] (2) 文書インデックス修正表示方法
 一般に、自動登録では、インデックス記載領域の文字パターンを認識してテキストデータに変換し、光ディスクに検索用キーワードとして書き込む。しかし、文字を100%正しく認識することは非常に困難であるので、精度の高い検索を行なうためには、認識結果を修正しておく必要がある。修正は、原画像と認識結果とを同時に表示し、両者を見比べながら、誤認識文字については、キーボードから正しい文字を入力することにより行なうが、登録する文書数が膨大なため、修正作業を文書管理部門以外の専門部に依頼する場合が多く、守秘項目はマスクして画像表示する必要が生じている。

[0035] 図2により、文書インデックス (検索用キーワード) 修正のための、画像表示方法を説明する。同図で、キーボードKBから「修正開始」要求を入力した後、

①手順THINで、修正者はKBから類似度閾値 Lth を入力する。これは、認識誤りの可能性が大きいインデックスほど、類似度が低いという事実を利用して、修正が必要な文書を選定するためである。

[0036] ②次に、手順SRHで、光ディスクのキーワードフィールド (図10、 K_i) から、類似度 lik_{ij} を読み出し、


[0037]


[数5]

$$\dots (数5)$$

$ij = 1$ なる領域 (守秘項目が記載されている領域) の座標データ sec_{ij} を光ディスクODから読み出し、該領域部分のみ "0" の値を持つマスクパターンを生成す

る。

【0039】④手順MSKおよびDSPで、ODから文書全体の画像データを読み出して画像処理プロセッサで伸張し、上記マスクパターンを重ねてディスプレイに表示する。

【0040】具体的には、を伸張後の原画像、*

$$Dspj = \text{imgj} \wedge Mij$$

… (数6)

を演算し(∧は論理積演算子を表わす)、 $Dspj$ を表示する。また、守秘領域が複数ある場合には、*

$$Dspj = ((\text{imgj} \wedge M1j) \wedge M2j) \wedge M3j \dots \dots (数7)$$

を表示する。

【0043】以上の手順による表示画面例を、図4に示す。図で、CRTはディスプレイ画面を、WINは画像表示用のウィンドウを示す。DOCは数6の演算による文書画像 $Dspj$ を表示した結果である。領域F3に守秘項目が記載されている場合で、F3の領域をマスクし、文字f3を覆い隠して表示している。

【0044】ここでは省略したが、別なウィンドウには、項目f1、f2、f4、およびfnをOCRで認識した結果(indij)と、KBから入力した修正文字列を表示する。

【0045】⑤図2手順REVでは、修正者は図4の画面を見ながら認識誤りがある場合は文書インデックスを修正する。手順ENTでは、修正したインデックスを再びODのキーワードフィールドに書き込む(旧インデックスを書き替える)。

【0046】⑥手順JINT、JEND、およびJIN★

$$\text{indij} = \text{TM}$$

なる文書を検索する。

【0051】④手順PCHKでは、PWINで入力されたパスワードPWと、あらかじめ電子ファイリング装置に検索者登録したパスワードPW_oとを照合し、それらが一致すれば手順IMG、手順DSPにより文書画像表示する。この手順による表示画面を図5に示す。同図で、文書画像DOCは、守秘領域をマスクせずに、全領域を表示している。

【0052】PW≠PW_oの場合は、手順PAT、MSK、およびDSPにより、図4と同様に守秘領域をマスクした表示を行なう。方法は、前述した修正表示と同一である。

【0053】⑥手順SENDで、検索終了がKBから入力されない場合は、手順TWINに戻って、次の文書検索処理を行なう。

【0054】なお、以上の実施例においては、電子ファイリング装置に自動登録して、修正検索する文書として、報酬明細書を取り挙げ、書式(項目記載領域の位置、大きさ等)が少しずつ相違している文書例を説明したが、書式情報は登録する文書に対応して作成するので、表形式文書であるならば報酬明細書、履歴書、勤務表等、様式の相違する文書が混在してもよい。

*Mijをマスクパターン(secij、すなわち(Xim, Yim)、ΔXim、およびΔYimで規定される範囲はMij=0、以外の範囲ではMij=1)として、

【0041】

【数6】

※【0042】

【数7】

★Cにより、 $likij \leq Lth$ なる文書がODに存在しなくなるまで、SRHからENTまでの一連の修正処理を繰り返す。

【0047】(3) 文書検索表示方法

図3により、文書検索での画像表示方法を説明する。同図で、キーボードKBから「検索開始」要求を入力した後、

①手順PWINで、使用者はKBからパスワードPWを入力する。

【0048】②手順TMINで、使用者はKBから検索ターム(検索条件)TMを入力する。

【0049】③次に、手順SRHでは、光ディスクのキーワードフィールド(図10でK1、K2、…、Kn)を読み出し、検索タームTMに一致するインデックスを持つ文書、すなわち、

【0050】

【数8】

… (数8)

【0055】また、メインメモリに格納し、画像マスクに用いる守秘項目記載領域の座標値は、書式情報ファイルの値ではなく、文書構造解析値(図1で説明した、手順②b)における((Xk, Yk)、ΔXk、ΔYk))であってもよい。さらに、書式情報ファイルの値による領域と、文書構造解析値による領域の和であってもよい。

【0056】

【発明の効果】以上述べた如く、本発明によれば、電子ファイリング装置において表形式文書を自動登録し、修正表示、および検索する上で、

①登録においては、文書構造解析で抽出した守秘項目記載領域の座標値を、画像データおよび文字認識部で認識したインデックスとともに光ディスクに蓄積し、

②修正・検索では、光ディスクから読み出した画像データの守秘項目記載領域は上記座標値にもとづいてマスクし、該領域以外は原画像の状態を表示するので、文書書式が相違しても、守秘項目記載領域のみをマスクして表示できるという効果を奏するものである。

【図面の簡単な説明】

【図1】本発明の実施例における文書画像の登録方法を示す図。

50 【図2】実施例における文書インデックス修正のための

画像表示方法を示す図。

【図3】実施例における文書検索での画像表示方法を示す図。

【図4】実施例における守秘領域をマスクした文書画像の表示画面を示す図。

【図5】実施例における文書画像の全領域を表示した画面を示す図。

【図6】実施例における書式を定量的に記述した、書式情報を示す図。

【図7】実施例における書式情報ファイルを示す図。

【図8】電子ファイリング装置で自動登録の対象とする文書の書式例を示す図。

【図9】本発明が適用される文書自動登録機能を有する電子ファイリング装置の構成を示す図。

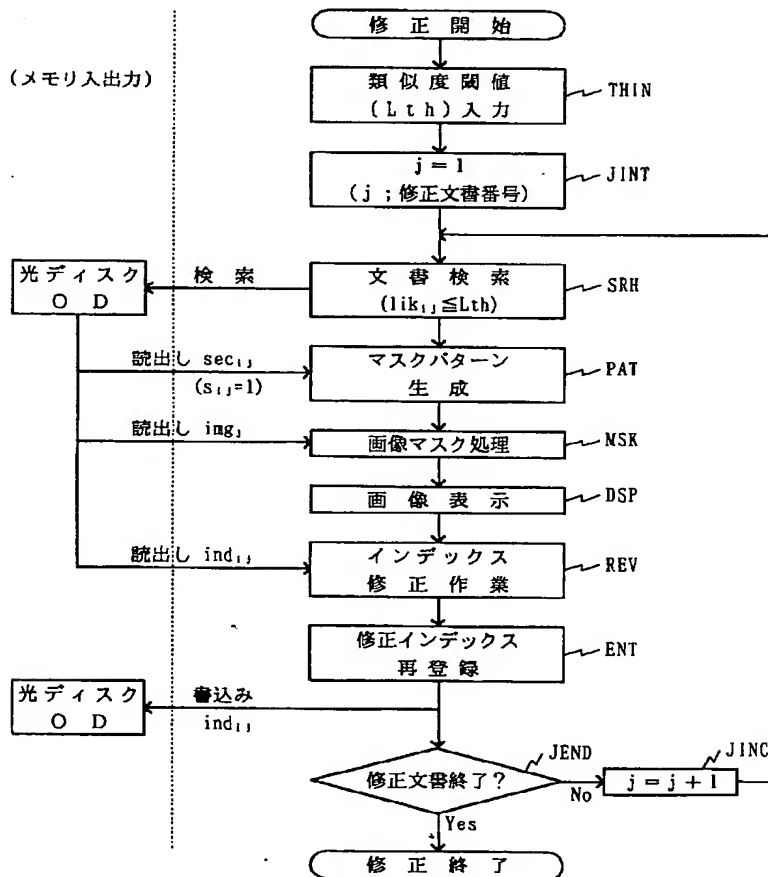
【図10】本発明が適用される電子ファイリング装置での光ディスクのキーワード情報の構成を示す図。

【符号の説明】

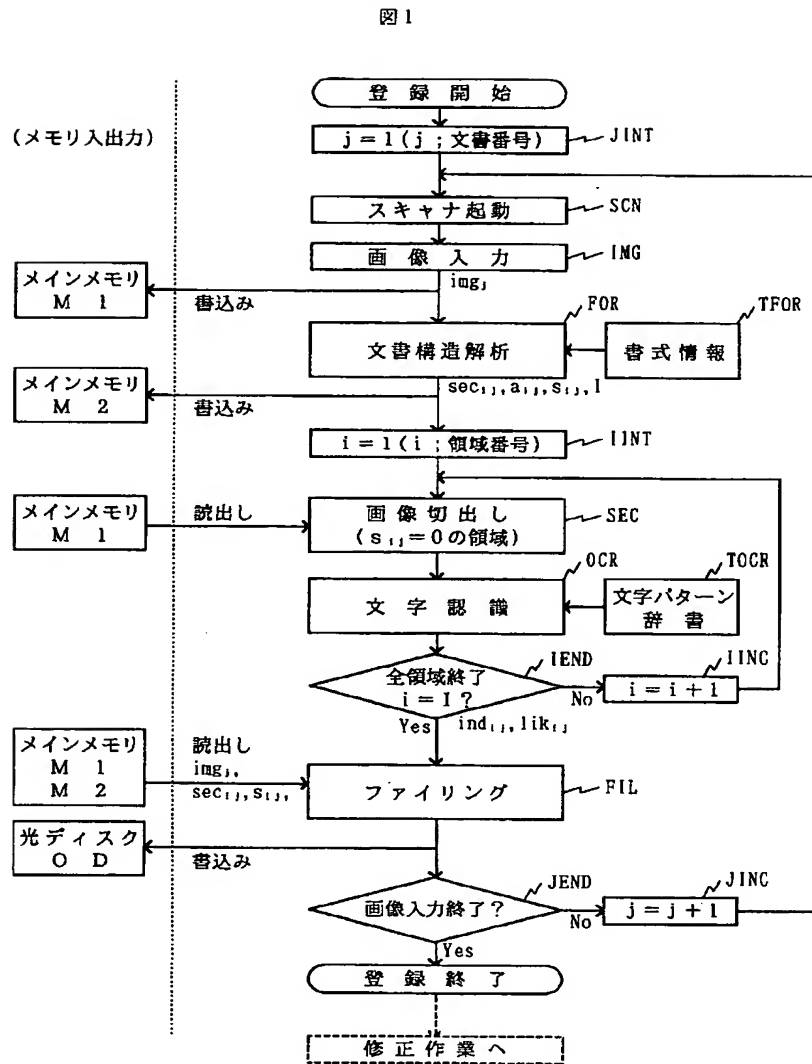
SCN…スキャナ、FOR…文書構造解析部、OCR…文字認識部、IPU…画像処理プロセッサ、OD…光ディスク、DK…磁気ディスク、MM…メインメモリ、 μ P…マイクロプロセッサ、Fi…インデックスまたは守秘項目記載領域、fi…インデックスまたは守秘項目、secij…インデックスまたは守秘項目記載領域座標データ、aij…インデックス記載領域の属性、sij…守秘項目識別情報、imgj…文書画像データ、indij…文字認識したインデックス、likij…文字認識の類似度、DOC…文書画像、MSK…画像マスク、TM…検索ターム、Ki…キーワード。

【図2】

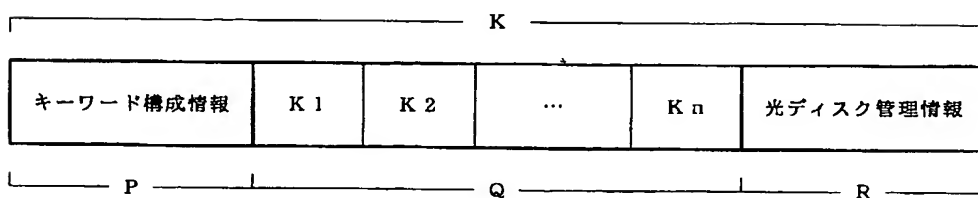
図2



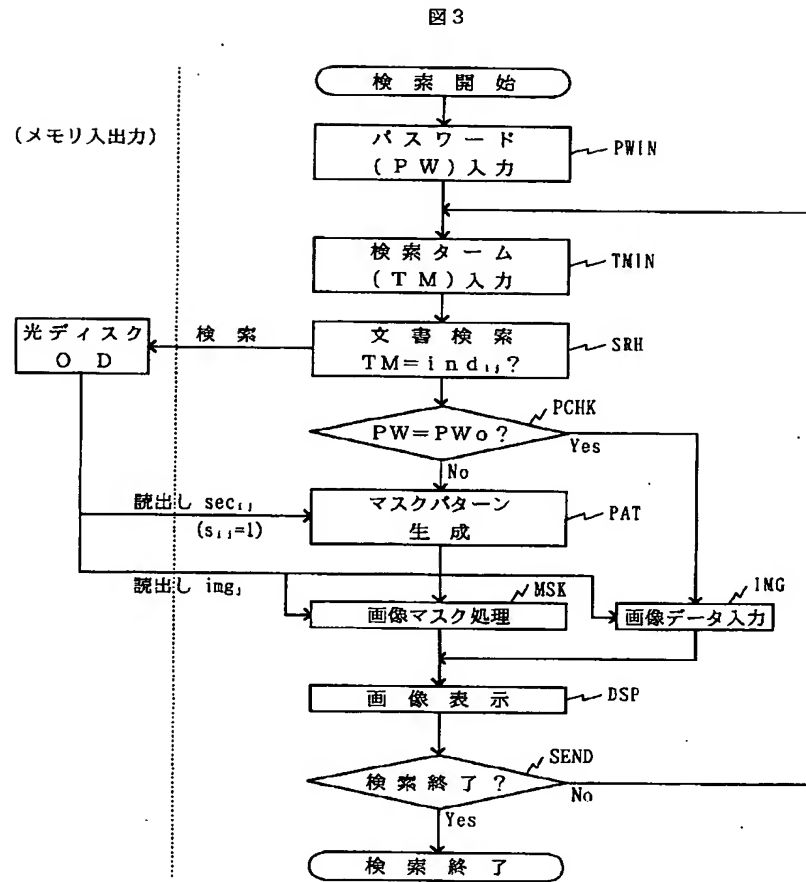
【図1】



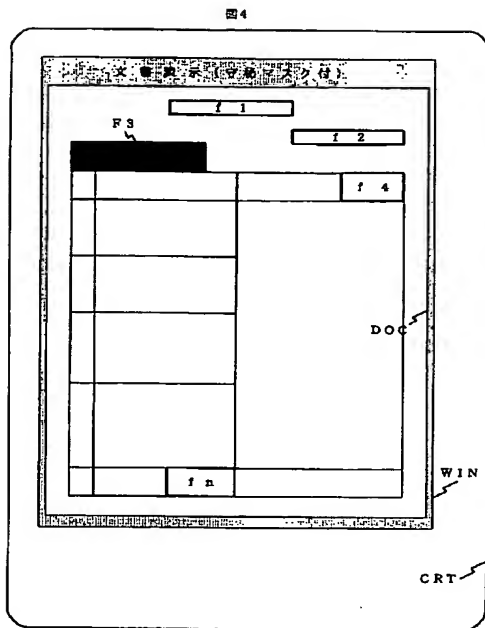
【図10】



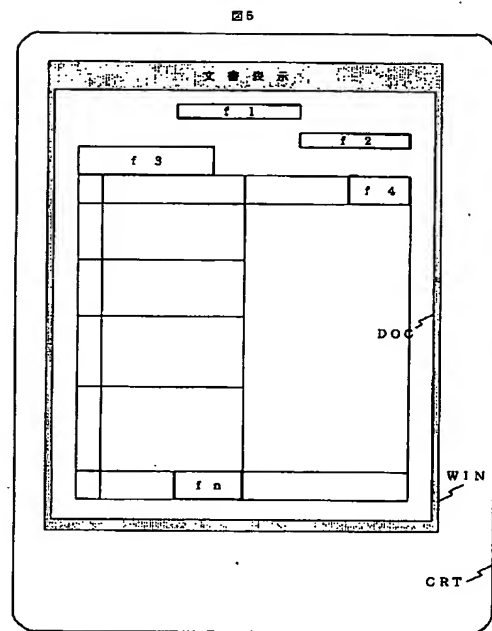
【図3】



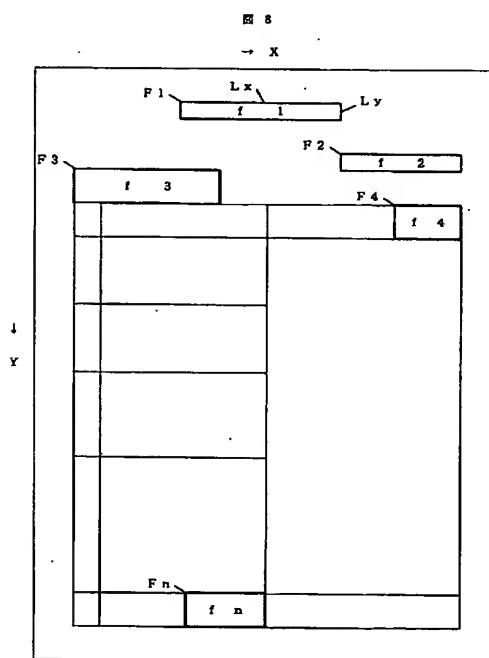
【図4】



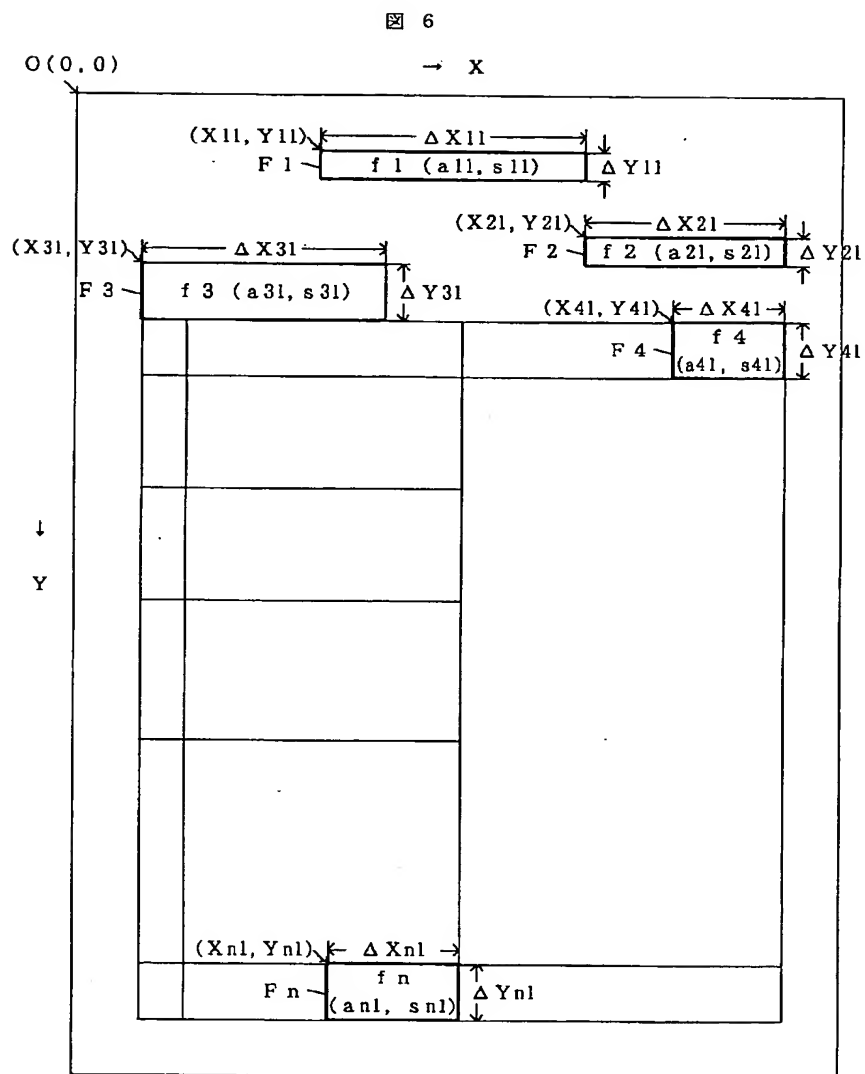
【図5】



【図8】



【図6】

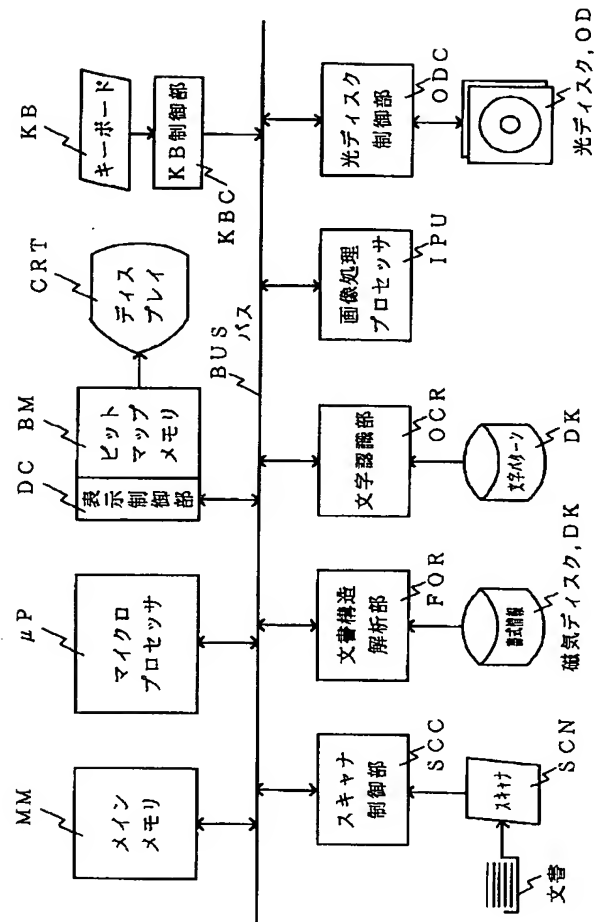


【図7】

図7

ADR					
1	(X11, Y11), ΔX11, ΔY11, a11, s11	← F 1	書式 1		
2	(X21, Y21), ΔX21, ΔY21, a21, s21	← F 2			
3	(X31, Y31), ΔX31, ΔY31, a31, s31	← F 3			
⋮	⋮				
n	(Xn1, Yn1), ΔXn1, ΔYn1, an1, sn1	← F n	書式 2		
n+1	(X12, Y12), ΔX12, ΔY12, a12, s12				
n+2	(X22, Y22), ΔX22, ΔY22, a22, s22				
n+3	(X32, Y32), ΔX32, ΔY32, a32, s32				
⋮	⋮				
2n	(Xn2, Yn2), ΔXn2, ΔYn2, an2, sn2				
⋮	⋮				
(m-1)n+1	(X1m, Y1m), ΔX1m, ΔY1m, a1m, s1m		書式 m		
(m-1)n+2	(X2m, Y2m), ΔX2m, ΔY2m, a2m, s2m				
(m-1)n+3	(X3m, Y3m), ΔX3m, ΔY3m, a3m, s3m				
⋮	⋮				
mn	(Xnm, Ynm), ΔXnm, ΔYnm, anm, snm				

【図9】



フロントページの続き

(72)発明者 羽田野 英一
東京都国分寺市東恋ヶ窪1丁目280番地
株式会社日立製作所中央研究所内

(72)発明者 藤縄 雅章
神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内